**Poster I-53**

## Programs for the Induction of Gene Networks and Their Analysis

*Livingston, Gary[*], Li, Guangyi, Hao, Liwu, Li, Xiao*
*University of Massachusetts at Lowell, Lowell, MA, USA*

The induction and analysis of gene networks generated from microarray gene expression data may greatly aid the understanding and cure of gene-based diseases such as cancer. We have devised a novel set of programs for (1) inducing gene networks from gene expression data that are based upon rule induction and (2) analyzing the induced networks. The first program, InduceNet, induces networks from gene expression data; CompareNet performs an edgewise comparison of two networks; and ComparePop uses a given network to compare the "fit" of the network to two populations. We demonstrate the utility of our programs by using them to analyze adenocarcinoma data published by Beer et al.

InduceNet uses a discovery program called HAMB to generate rules predicting the expression levels of each gene by using the expression levels of the other genes. Rule sets using one gene to predict another with an accuracy (proportion of correct predictions) exceeding a given threshold are used to form the edges of the induced network.

CompareNet creates a "difference network" from two networks by comparing the edges in the two networks, thus creating a new network comprised of the edges that differ between the networks as follows: edges that appear in the first network but not in the second, edges that appear in the second network but not in the first, and edges that appear in both networks, but with different correlations (e.g., a positive correlation in the first network and a negative correlation in the second).

ComparePop compares two populations using a given network. For each edge in the given network, ComparePop computes p-values for the differences. The p-values are calculated from the proportions of the two populations which are correctly predicted by rule sets created from the edges. A threshold of the p-values is used to identify the edges which significantly differ with respect to their fit with the given network.

CompareNet and ComparePop are not limited to analyzing networks generated by InduceNet; they may be used to analyze any gene networks. For example, CompareNet could be used to compare an induced network to a manually derived network, such as a known gene network. Thus, CompareNet could be used to compare a network induced from cancer data to a manually generated cancer network.

Our current research focus is to extend our program to use ontological data. We are developing methods for using ontological data and models of gene expression and cancer to automatically evaluate and explain the results of our programs and propose hypotheses for surprising findings.

*Figure 1. A portion of a graph generated by ComparePop, using a gene network induced by InduceNet from the Beer et al. data set to identify differences between Stage 3 tumor cases and nontumor cases.*